

## Chord spacing and quality: Lessons from timbre research

Noah Kahrs<sup>1†</sup>

Matt Chiu<sup>1</sup>

<sup>1</sup> Eastman School of Music, University of Rochester, Rochester, NY, USA

<sup>†</sup> Corresponding author: [nkahrs@u.rochester.edu](mailto:nkahrs@u.rochester.edu)

Published 16 December 2021; <https://doi.org/10.18061/FDMC.2021.0050>

Author video presentation and/or other conference material: <https://doi.org/10.17605/OSF.IO/G9DYF>

### Abstract

Although chords are often represented by pitch-class (chroma) content in computational research, chord spacing is often a more salient feature. This paper addresses this disparity between models and cognition by extending the discrete Fourier transform (DFT) theory of chord quality from pitch-classes to pitches. In doing so, we note a structural similarity between music theory's chord quality and audio engineering's timbral cepstrum: both are DFTs, performed in the pitch or frequency domains, respectively. We thus treat chord spacing as a hybrid of pitch-class and timbre.

To investigate the potential benefits of the DFT on pitch space (P-DFT), we perform two computational experiments. The first explores the P-DFT model theoretically by correlating chord distances calculated with a pitch-class model against those calculated with spacing. The second compares P-DFT estimations of chord distances against listener responses (Kuusi, 2005). Our results show that spacing is a salient feature of chords, and that it can be productively described by timbre-influenced methods.

**KEYWORDS:** *chord spacing, chord quality, timbre, Discrete Fourier Transform, cepstrum*

### Introduction

Whether considering chords (Burgoyne et al., 2011), keys (Albrecht & Shanahan, 2013; Temperley & Marvin, 2008), or musical style (Yust, 2019), computational and corpus models of music tend to rely on pitch-classes (PCs, also known as pitch chroma), not pitches or frequencies. Although the assumption of octave equivalence may seem innocuous, given the crucial role of pitch chroma to perception (Krumhansl, 1979), listeners often identify chord similarity more by spacing than by PC content (Kuusi, 2005; Samplaski, 2001). To consider the role of spacing in chord quality, this paper takes advantage of a conceptual similarity between uses of the Discrete Fourier Transform (DFT) on pitch-classes in music theory and on spectra in computational timbre research.

The DFT is an equation that breaks a signal into periodic components. A DFT on a signal returns a set of

values  $f_0-f_n:f_n$  (component  $n$ ) represents a division of the signal's underlying space into  $n$  equal parts. If a signal approximates an even division, then the *magnitude* for that component will be high. Figure 1 shows how DFTs are used in different disciplines. The DFT can be applied to a waveform to yield the pitch spectrum, which is commonly used in timbre research to yield cognitively salient timbre descriptors (Peeters et al., 2011). Shown on the same level, pitch and pitch-class content as used in music theory and pitch-based music cognition are conceptually equivalent in that they both use pitch data.

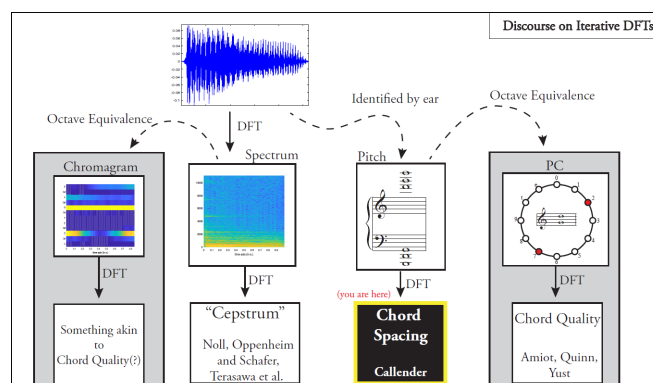


Figure 1: Summary of Previous DFT Research

Our paper introduces a hitherto unnoticed similarity between the cepstrum—a DFT on the spectrum's DFT data—and music theory's "chord quality," a DFT on pitch-class data (henceforth PC-DFT). The PC-DFT is used by mathematical music theorists to describe chord quality, which emerges from periodicities in pitch-class space (Amiot, 2016; Chiu, in press; Quinn, 2007; Yust, 2015). The cepstrum, the DFT of the spectrum, is also derived from periodicities in pitch data, albeit on all frequencies from the waveform rather than only on fundamentals. The cepstrum is used most often for timbre identification, (Aucouturier & Pachet, 2004; Herrera-Boyer et al., 2006), and despite its reputation for cognitive opacity, it does correlate with listener judgments of timbre similarity (Aucouturier & Bigand,



2012; Casey et al., 2012; Siedenburg et al., 2016; Terasawa et al., 2005).

Our P-DFT fuses these approaches by considering notated pitches, like the PC-DFT, but avoiding octave equivalence. Such a model was previously proposed by Callender (2007) in a continuous space; we expand on his work with both a more detailed examination of the P-DFT's relation to chord spacing, and a comparison to listener models of chord similarity. We thus study chord spacing using a combination of features from research in both pitch and timbre cognition, continuing in a tradition of investigating pitch-timbre relationships (Allen & Oxenham, 2014; Hasegawa, 2019; Krumhansl & Iverson, 1992; Saariaho, 1987).

To that end, we performed two exploratory computational studies to investigate how the P-DFT's spacing-oriented assessment of chord similarity compares to other measures, in both music-analytical and experimental contexts.

### General Method: Calculating and Interpreting P-DFT Components

Both experiments used the same procedure to calculate chords' P-DFTs. First, chords were encoded as normalized characteristic functions in MIDI pitch space. For example, a closed-position C-major chord beginning on middle C would consist of notes C4, E4, and G4, or MIDI values 60, 64, and 67. The corresponding characteristic function would be a 128-place vector with the value .33 at positions 60, 64, and 67, and 0 at all other positions. Then, the vectors were zero-padded (128 more zeros were appended) to avoid wrap-around effects.

A DFT was calculated on this vector of 256 values, yielding a new vector of 256 component magnitudes. Due to the Nyquist effect, only the first 128 of these components were non-redundant; the rest were discarded. Distances between chords were calculated as Euclidean distances (Albrecht & Shanahan, 2013; Callender, 2007) between P-DFT vectors normalized in magnitude.

The  $n$ th component corresponds to chord spacing of  $256/n$  semitones. For example, an octave is 12 semitones, and  $256/12 = 21\frac{1}{3}$ , so the corresponding components for the octave—21 and 22—would be high. Because this relationship is inversely proportional, low components correspond to wide spacing, whereas high components indicate specific smaller intervals. This parallels a similar situation in the cepstrum, in which only low components are used for timbre identification,

whereas higher ones are more useful for pitch identification (Aucouturier & Pachet, 2004; Muller & Ewert, 2010).

Figure 2 shows the P-DFT component magnitudes for the iconic opening chord of Claude Debussy's "Sunken Cathedral" Prelude (henceforth the "Cathedral Chord"). We have circled peaks corresponding to the four-octave space between the two hands, as well as the open spacing, captured by peaks at the fourth, fifth, and tritone average within each hand.

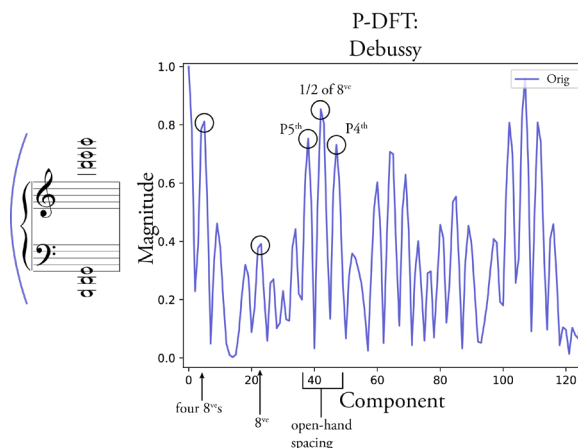


Figure 2: Example DFT Calculation

### Experiment 1: Similar Chords to Debussy's "Sunken Cathedral" under P- and PC-DFT

#### Method

To determine what sorts of musical features are weighed by the P-DFT vs by the PC-DFT, we compared the opening chord of Debussy's "Sunken Cathedral" prelude to 50,000 random chords.

Chords were randomly generated by first selecting the number of pitches to be included (2 to 6), then independently selecting that many notes from the range of a standard grand piano (MIDI values 21 to 108) and removing duplicates. Distances were calculated from each random chord to Debussy's original chord as described above.

We hypothesized that chords similar to (that is, low distance from) Debussy's original chord by PC-DFT would be of a similar PC set-class (that is, related by transposition or inversion accounting for octave equivalence; see Morris, 1987), and that chords similar by P-DFT would have similar spacing on the piano keyboard.

## Results

Figure 3 plots dissimilarity of P- and PC-DFTs of all random chords, as measured against the Cathedral Chord. Although P- and PC-DFTs are correlated ( $r = .057$ ,  $p < .001$ ), there are clearly outliers. By investigating two outliers (selected by their maximal distance from the line extending from minimum to maximum P- and PC-DFT distances), we can see which features most directly affect assessed similarity.

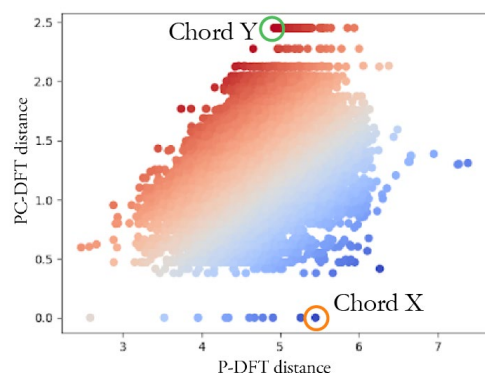


Figure 3: PC- vs P-DFT distances of random chords versus Cathedral Chord

Figure 4 compares the P-DFTs of the Cathedral Chord (Orig) and Chord X. Chord X is identical in PC content, and thus has PC-DFT distance 0. However, as an open eleventh, it has a completely different spacing, and thus its P-DFT has completely different components.

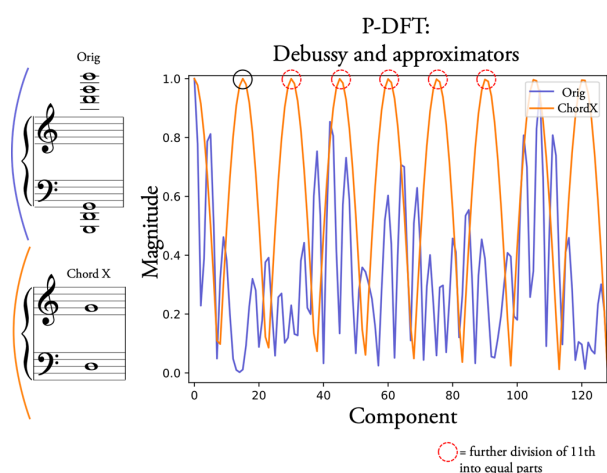


Figure 4: Cathedral Chord vs PC-similar chord

Figure 5 compares the Cathedral Chord to Chord Y, which is dissimilar in PC-DFT space, but somewhat similar in P-DFT space. It belongs to set-class (013579), which is almost a whole-tone scale and is very different from open fifths or fourths. However, its spacing is quite

similar to that of the Cathedral Chord, as shown by shared peaks at components corresponding to wide spacing, octaves, and division of octaves into approximate halves (fifths/sixths).

## Discussion

As we predicted, the P-DFT measures some aspect of chord spacing consistent with a subjective impression of Debussy's prelude; the prelude's opening, noted for its

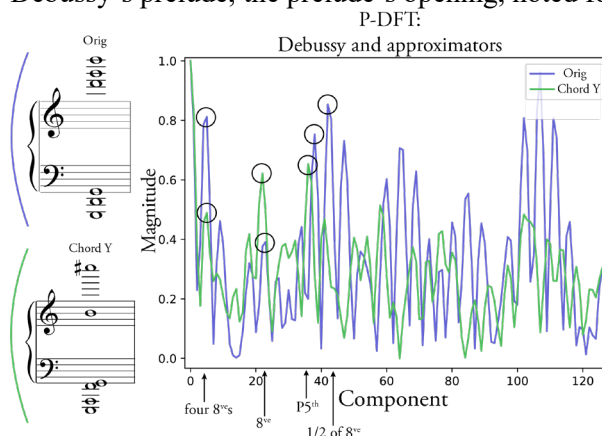


Figure 5: Cathedral Chord vs P-similar chord

wide spacing and “deep bass chords,” (Roberts, 2001, p. 261) is reminiscent of a bell and the harmonic series. Such features are absent from the simple open eleventh of Chord X, but present in Chord Y despite its very different PC content.

These findings show that the P-DFT captures a musically relevant aspect of a chord's quality. Because the Cathedral Chord and Chord Y shared several similar peaks among low components, the P-DFT specifically captures wide-interval spacing. As low components measure timbre in the cepstrum, these components show that the P-DFT's chord spacing is sonically similar to timbre.

## Experiment 2: Correlations of P-DFT Similarity with Listener Judgments

### Method

In order to verify the model's cognitive salience, we compared P-DFT components with perceptual similarity ratings for chord distances from Kuusi (2005).

Kuusi's experiment explored correlations between subject responses and set-class similarity measures in a post-tonal setting. Her stimuli consisted of 16 chords: 4 different set classes each with 4 different spacings. Subjects were asked to rate the similarity of all possible

pairs of chords on a Likert scale of 1–7. Subject ratings were correlated with set-class similarity measures; the measure yielding the highest correlation with subject responses was CSATSIM (Buchler, 1997) ( $r = .43$ ). Kuusi’s subjects included both trained and less-trained musicians; we considered only the data from trained musicians. In a post-hoc analysis, Kuusi found that the subject-rated similarity between two chords increased with the difference in chord span (distance between the outer two notes of the chord).

Seeing the impact of chord spacing, we hypothesized that the Euclidean Distance between chords’ P-DFT components would correlate more with subject data than set-class measures do. Such a finding would support our argument that the P-DFT, in measuring chord spacing, captures a cognitively meaningful aspect of chords erased by octave equivalence.

### Results

The Pearson correlation between P-DFT distances and subject distances were significantly correlated ( $r = .38$ ,  $p < .001$ ), but less so than CSATSIM. However, a stepwise regression yielded a subset of 49 components (1–5, 7–18, 21, 25, 30, 45, 61–66, 70, 84–88, 93, 96, 100, 105–110, 114, 122–126, 128) that was more correlated with subject responses ( $r = .74$ ,  $p < .001$ ).

### Discussion

As the P-DFT components describe different elements of chord spacing, we expected distances between P-DFTs to correlate with Kuusi’s subject ratings. However, distances between raw P-DFT calculations are still outperformed by some set-class measures, perhaps because subjects use asymmetrical weightings for chord distances: P-DFT distances compare all 128 components equally rather than weighting particular features, whereas listeners might prioritize certain intervals over others. To identify such intervals in the experiment, we isolated and correlated individual component similarities with listener ratings of chord similarity.

Figure 6 shows that the 49 components selected by stepwise regression tend to cluster where components are particularly correlated or uncorrelated with listener ratings. Out of the 49 components, 32% corresponded to spacing wider than an octave (between  $f_0$ – $f_{20}$ )—something that could not be captured by PC measures. Later components clustered around smaller intervals that listeners may have found crucial in chord identification, and which were likely accounted for in PC measures such as CSATSIM. Because this specific

combination of wide spacing and specific intervals better predicted correlations with listener data than PC similarity measures did ( $r = .74$ ), this result implies that our P-DFT model captures subjects’ dual reliance on general spacing and on specific interval content.

Kuusi’s experiment intended to produce a non-tonal environment, but, depending on the musical context, listeners may rely on different paradigms for evaluating chord proximity. Future experiments should explore this disparity, and the post-hoc exploration from Experiment 2 suggests that the P-DFT might be an effective tool for doing so. Furthermore, as listeners might prioritize certain intervals over others, future computational work might see if adjusted component weights better predicts listener responses, or if pitch-class measures provide additional non-redundant information.

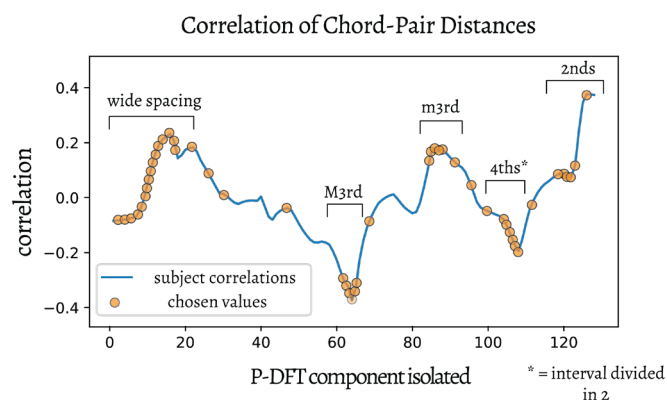


Figure 6. P-DFT Individual component correlations

### Conclusion

Both of our experiments show that the P-DFT captures aspects of chord spacing that are salient, whether in the historical and music-theoretical reception of a piece or when comparing chords in isolation. As our theoretical framework demonstrates a formal similarity between chord quality and timbre similarity, our results fit with the growing literature on interactions between pitch, harmony, and timbre. Furthermore, by associating certain P-DFT components with specific chord spacings as well as listener similarity, we suggest that cepstrum-like methods may in fact be of potential use in cognitive studies.

### Acknowledgements

We would like to thank Tuire Kuusi for providing us with data from her 2005 study, and David Temperley for providing comments on a draft of this article.



## References

- Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception*, 31(1), 59–67. <https://doi.org/10.1525/mp.2013.31.1.59>
- Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, 135(3), 1371–1379. <https://doi.org/10.1121/1.4863269>
- Amiot, E. (2016). *Music through Fourier space*. Springer. <https://doi.org/10.1007/978-3-319-45581-5>
- Aucouturier, J., & Pachet, F. (2004). Improving Timbre Similarity: How high's the sky? *Journal of negative results in speech and audio sciences*, 1(1).
- Aucouturier, J.-J., & Bigand, E. (2013). Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, 41(3), 483–497. <https://doi.org/10.1007/s10844-013-0251-x>
- Buchler, M. H. (1997). *Relative saturation of subsets and interval cycles as a means for determining set-class similarity* [Doctoral dissertation, University of Rochester]. ProQuest Dissertations Publishing.
- Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis. *ISMIR*, 11, 633–638.
- Callender, C. (2007). Continuous harmonic spaces. *Journal of Music Theory*, 51(2), 277–332. <https://doi.org/10.1215/00222909-2009-004>
- Casey, M., Thompson, J., Kang, O., Raizada, R., & Wheatley, T. (2012). Population codes representing musical timbre for high-level fMRI categorization of music genres. In Georg Langs, Irina Rish, Moritz Grosse-Wentrup, & Brian Murphy (Eds.), *Machine Learning and Interpretation in Neuroimaging* (pp. 34–41). Springer. [https://doi.org/10.1007/978-3-642-34713-9\\_5](https://doi.org/10.1007/978-3-642-34713-9_5)
- Chiu, M. (in press). Macroharmonic Progressions through the Discrete Fourier Transform: An Analysis from Maurice Duruflé's Requiem. *Music Theory Online*, 27(3).
- Hasegawa, R. (2019). Timbre as Harmony—Harmony as Timbre. In E. I. Dolan & A. Rehding (Eds.), *The Oxford Handbook of Timbre*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190637224.013.11>
- Herrera-Boyer, P., Klapuri, A., & Davy, M. (2006). Automatic classification of pitched musical instrument sounds. In *Signal processing methods for music transcription* (pp. 163–200). Springer. [https://doi.org/10.1007/0-387-32845-9\\_6](https://doi.org/10.1007/0-387-32845-9_6)
- Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive psychology*, 11(3), 346–374. [https://doi.org/10.1016/0010-0285\(79\)90016-1](https://doi.org/10.1016/0010-0285(79)90016-1)
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 739. <https://doi.org/10.1037/0096-1523.18.3.739>
- Kuusi, T. (2005). Chord span and other chordal characteristics affecting connections between perceived closeness and set-class similarity. *Journal of New Music Research*, 34(3), 259–271. <https://doi.org/10.1080/09298210500236325>
- Morris, R. (1987). *Composition with pitch-classes*. Yale University Press. <https://doi.org/10.2307/j.ctt1xp3ss4>
- Muller, M., & Ewert, S. (2010). Towards Timbre-Invariant Audio Features for Harmony-Based Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3), 649–662. <https://doi.org/10.1109/TASL.2010.2041394>
- Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902–2916. <https://doi.org/10.1121/1.3642604>
- Quinn, I. (2007). General equal-tempered harmony: Parts 2 and 3. *Perspectives of New Music*, 4–63. <https://doi.org/10.1353/pnm.2007.0016>
- Roberts, P. (2001). *Images: the piano music of Claude Debussy*. Hal Leonard Corporation.
- Saariaho, K. (1987). Timbre and harmony: Interpolations of timbral structures. *Contemporary Music Review*, 2(1), 93–133. <https://doi.org/10.1080/07494468708567055>
- Samplaski, A. G. (2000). *A comparison of perceived chord similarity and predictions of selected twentieth-century chord-classification schemes, using multidimensional scaling and cluster analysis*. [Doctoral dissertation, Indiana University]. ProQuest Dissertations Publishing.
- Siedenburg, K., Fujinaga, I., & McAdams, S. (2016). A Comparison of Approaches to Timbre Descriptors in Music Information Retrieval and Music Psychology. *Journal of New Music Research*, 45(1), 27–41. <https://doi.org/10.1080/09298215.2015.1132737>
- Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception*, 25(3), 193–212. <https://doi.org/10.1525/mp.2008.25.3.193>
- Terasawa, H., Slaney, M., & Berger, J. (2005). The thirteen colors of timbre. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*. (pp. 323–326). IEEE. <https://doi.org/10.1109/ASPAA.2005.1540234>
- Yust, J. (2015). Schubert's harmonic language and Fourier phase space. *Journal of Music Theory*, 59(1), 121–181. <https://doi.org/10.1215/00222909-2863409>

Yust, J. (2019). Stylistic information in pitch-class distributions. *Journal of New Music Research*, 48(3), 217–231. <https://doi.org/10.1080/09298215.2019.1606833>